Attorney Docket No. 02307O-086120

PATENT APPLICATION

NUCLEIC ACIDS THAT CONTROL SEED AND FRUIT DEVELOPMENT IN PLANTS

Inventors:

Robert L. Fischer
Nir Ohad
Tomohiro Kiyosue
Ramin Yadegari
Linda Margossian
John Harada
Robert B. Goldberg

Assignee:

Regents of the University of California

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 576-0200

CROSS REFERENCE TO RELATED APPLICATIONS

This is a continuation-in-part of USSN 09/071,838, filed May 1, 1998, which is incorporated herein by reference.

FIELD OF THE INVENTION

The present invention is directed to plant genetic engineering. In particular, it relates to modulation of expression of genes controlling endosperm development in plants.

BACKGROUND OF THE INVENTION

A fundamental problem in biology is to understand how fertilization initiates reproductive development. In higher plants, the ovule generates the female gametophyte which is composed of egg, central, synergid and antipodal cells (Reiser, *et al., Plant Cell*, 1291-1301 (1993)). All are haploid except the central cell which contains two daughter nuclei that fuse prior to fertilization. One sperm nucleus fertilizes the egg to form the zygote, whereas another sperm nucleus fuses with the diploid central cell nucleus to form the triploid endosperm nucleus (van Went, *et al., Embryology of Angiosperms*, pp. 273-318 (1984)). The two fertilization products undergo distinct patterns of development. In Arabidopsis, the embryo passes through a series of stages that have been defined morphologically as preglobular, globular, heart, cotyledon and maturation (Goldberg, R. B., *et al., Science* (1994) 266: 605-614; Mansfield, S. G., *et al., Arabidopsis: An Atlas of Morphology and Development*, pp. 367-383 (1994)). The primary endosperm nucleus undergoes a series of mitotic divisions to produce nuclei that migrate into the expanding central cell (Mansfield, S. G., *et al., Arab Inf Serv* 27: 53-64 (1990); Webb, M. C., *et al., Planta* 184: 187-195 (1991)). Cytokinesis sequesters endosperm cytoplasm and nuclei into discrete cells (Mansfield, S. G., *et al., Arab Inf Serv* 27: 65-72 (1990)) that produce storage proteins, starch, and lipids which support embryo growth (Lopes, M. A. *et al., Plant Cell* 5: 1383-1399 (1993)). Fertilization also activates development of the integument cell layers of the ovule that become the seed coat, and induces the ovary to grow and form the fruit, or silique, in Arabidopsis.

Control of the expression of genes that control egg and central cell differentiation, or those that activate reproductive development in response to fertilization is

useful in the production of plants with a range of desired traits. These and other advantages are provided by the present application.

SUMMARY OF THE INVENTION

The present invention provides methods of modulating fruit and seed development and other traits in plants. The methods involve providing a plant comprising a recombinant expression cassette containing an *FIE* nucleic acid linked to a plant promoter.

In some embodiments, transcription of the *FIE* nucleic acid inhibits expression of an endogenous *FIE* gene or activity the encoded protein. This embodiment is particularly useful, for instance, making embryo-less seed and parthenocarpic fruit. Alternatively, expression of the *FIE* nucleic acid may enhance expression of an endogenous *FIE* gene or FIE activity

In the expression cassettes, the plant promoter may be a constitutive promoter, for example, the CaMV 35S promoter. Alternatively, the promoter may be a tissue-specific promoter. Examples of tissue specific expression useful in the invention include ovule-specific or embryo-specific expression. For instance, the promoter sequence from the *FIE* genes disclosed here can be used to direct expression in relevant plant tissues.

The invention also provides seed or fruit produced by the methods described above. The seed or fruit of the invention comprise a recombinant expression cassette containing an *FIE* nucleic acid.

Definitions

The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role..

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants. Promoters include necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and

developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The term "plant" includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, and the like. Such a plant containing the exogenous nucleic acid is referred to here as an R_1 generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are descendants of such a plant.

A "*FIE* nucleic acid" or "*FIE* polynucleotide sequence" of the invention is a subsequence or full length polynucleotide sequence of a gene which encodes a polypeptide involved in control of reproductive development and which, when mutated, allows for aspects of fertilization independent reproductive development. In some embodiments, the polypeptides of the invention have substantial sequence identity (as defined below) to a polycomb group gene of *Drosophila*. An exemplary nucleic acid of the invention is the *Arabidopsis FIE1* and *FIE3* sequences disclosed below. *FIE* polynucleotides are defined by their ability to hybridize under defined conditions to the exemplified nucleic acids or PCR

products derived from them. An *FIE* polynucleotide is typically at least about 30-40 nucleotides to about 3000, usually less than about 5000 nucleotides in length. The nucleic acids contain coding sequence of from about 100 to about 2000 nucleotides, often from about 500 to about 1700 nucleotides in length.

FIE nucleic acids are a new class of plant regulatory genes that encode polypeptides with sequence identity to members of the polycomb group genes first identified in *Drosophila*. Polycomb group gene products and their homologues in other species are responsible for repression of homeotic genes. The proteins are a heterogenous group that interact with each other to form large complexes that bind DNA and thereby control gene expression. For a review of the current understanding of polycomb complex genes *see*, Pirrotta *Cur. Op. Genet. Dev.* 7:249-258 (1997). Nine groups of polycomb genes have been identified. *FIE1* (SEQ ID NO:1) is related to the group of polycomb genes encoding protein comprising a SET domain (*see, e.g.*, Jenuwein *et al. Cell. Mol. Life Sci.* 54:80-93 (1998). *FIE3* (SEQ ID NO:3) is related to the group encoding proteins comprising WD40 repeats (*see*, Gutjahr *et al. EMBO J.* 14:4296-4306 (1995).

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only "substantially identical" to a sequence of the gene from which it was derived. As explained below, these substantially identical variants are specifically covered by the term *FIE* nucleic acid.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the terms "*FIE* nucleic acid". In addition, the term specifically includes those sequences substantially identical (determined as described below) with an *FIE* polynucleotide sequence disclosed here and that encode polypeptides that are either mutants of wild type FIE polypeptides or retain the function of the FIE polypeptide (*e.g.*, resulting from conservative substitutions of amino acids in the FIE polypeptide). In addition, variants can be those that encode dominant negative mutants as described below.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences,

refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, *Computer Applic. Biol. Sci.* 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA)..

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to sequences or subsequences that have at least 60%, preferably 80%, most preferably 90-95% nucleotide or amino acid residue identity when aligned for maximum correspondence over a comparison window as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. This definition also refers to the complement of a test sequence, which has substantial sequence or subsequence complementarity when the test sequence has substantial identity to a reference sequence.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The

sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al., J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the

codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(*see, e.g.*, Creighton, *Proteins* (1984)).

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical

is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30 °C below the T_m. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 time background hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cased, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

In the present invention, genomic DNA or cDNA comprising *FIE* nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B show the genetic map used to clone the *FIE3* gene.

Figure 2 shows the analysis of the sequence in the DNA shown in Figure 1 using the GENSCANW program.

Figure 3 shows the position of primers used to PCR amplify sequences from the *FIE3* gene region.

Figure 4 shows the genetic map used to clone the *FIE1* gene.

Figure 5 shows the results of complementation tests establishing that a single gene (*FIE1*) was present on the complementing cosmid (6-22) that was not fully encoded on either of the non-complementing cosmids (2-9 and 2-8).

DESCRIPTION OF THE PREFERRED EMBODIMENTS

This invention provides molecular strategies for controlling seed and fruit development.

Reproduction in higher plants is unique because it is initiated by two fertilization events in the haploid female gametophyte. One sperm nucleus fertilizes the egg to form the embryo. A second sperm nucleus fertilizes the central cell to form the endosperm, a unique tissue that supports the growth of the embryo. Fertilization also

activates maternal tissue differentiation, the ovule integuments form the seed coat and the ovary forms the fruit.

The present invention is based, at least in part, on the discovery of a set of female-gametophytic mutations, termed fie (fertilization-independent endosperm), and the subsequent cloning of the genes involved. Three mutants are disclosed here *fie1*, *fie2*, and *fie3*, which have been mapped to chromosomes 1, 2, and 3 of *Arabidopsis*, respectively. The fie mutations affect the central cell, allowing for replication of the central cell nucleus and endosperm development without fertilization. FIE/fie seed coat and fruit undergo fertilization-independent differentiation, showing that the fie female gametophyte is the source of signals that activates sporophytic fruit and seed coat development. Generally, the mutant *fie* alleles are not transmitted by the female gametophyte. Inheritance of a mutant *fie* allele (*e.g.*, *fie3*) by the female gametophyte usually results in embryo abortion, even when the pollen bears the wild-type *FIE* allele. In the case of *fie1* and *fie2*, however, transmission of the trait occurs in about 1% of the progeny from the female gametophyte. In contrast, the *fie1*, *fie2*, and *fie3* mutant alleles are passed through the male gametophyte (*i.e.*, pollen) in normal fashion.

The isolated sequences prepared as described herein, can be used in a number of techniques, for example, to suppress or enhance endogenous FIE gene expression. Modulation of *FIE* gene expression or FIE activity in plants is particularly useful, for example, in producing embryo-less seed, parthenocarpic fruit, or as part of a system to generate apomictic seed.

Isolation of *FIE* nucleic acids

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989).

The isolation of *FIE* nucleic acids may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can

be used to identify the desired gene in a cDNA or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ, such as ovules, and a cDNA library which contains the *FIE* gene transcript is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which *FIE* genes or homologs are expressed.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned *FIE* gene disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Alternatively, antibodies raised against an FIE polypeptide can be used to screen an mRNA expression library.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the *FIE* genes directly from genomic DNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. For a general overview of PCR see *PCR Protocols: A Guide to Methods and Applications*. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), *Academic Press*, San Diego (1990).

Appropriate primers and probes for identifying *FIE* sequences from plant tissues are generated from comparisons of the sequences provided here with other polycomb group genes. For instance, *FIE1* can be compared to the other polycomb genes containing the SET domain, such as the *Arabidopsis* curly leaf gene (Goodrich *et al. Nature* 386:44-51 (1997)) or the *Drosophila enhancer of zeste (E(z))* gene. *FIE3* can be compared to genes containing WD40 repeats, such as the *extra sex combs (esc)* gene from *Drosophila*. Using these techniques, one of skill can identify conserved regions in the nucleic acids disclosed here to prepare the appropriate primer and probe sequences. Primers that specifically hybridize to conserved regions in *FIE1* or *FIE3* genes can be used to amplify sequences from widely divergent plant species.

Standard nucleic acid hybridization techniques using the conditions disclosed above can then be used to identify full length cDNA or genomic clones.

Control of FIE activity or gene expression

Since *FIE* genes are involved in controlling seed, in particular endosperm, development, inhibition of endogenous Fie activity or gene expression is useful in a number of contexts. For instance, inhibition of expression is useful in the development of parthenocarpic fruit (*i.e.*, fruit formed in the absence of fertilization).

In addition, inhibition of FIE activity can be used for production of fruit with small and/or degraded seed (referred to here as "seedless fruit") after fertilization. In many plants, particularly dicots, the endosperm is not persistent and eventually is degraded. Thus, in plants of the invention in which Fie activity is inhibited, embryo-less seed do not persist and seedless fruit are produced.

Alternatively, plants of the invention can be used to prevent pre-harvest sprouting in seeds, especially those derived from cereals. In these plants, the endosperm persists and is the major component of the mature seed. Premature growth of embryos in stored grain causes release of degradative enzymes which digest starch and other components of the endosperm. Plants of the present invention are useful in addressing this problem because the seeds lack an embryo and thus will not germinate.

In yet another use, nucleic acids of the invention can be used in the development of apomictic plant lines (*i.e.*, plants in which asexual reproductive processes occur in the ovule, *see*, Koltunow, A. *Plant Cell* 5: 1425-1437 (1993) for a discussion of apomixis). Apomixis provides a novel means to select and fix complex heterozygous genotypes that cannot be easily maintained by traditional breeding. Thus, for instance, new hybrid lines with desired traits (*e.g.*, hybrid vigor) can be obtained and readily maintained.

In still another use, nucleic acids of the invention can be used to control endosperm production in transgenic plants. In particular, inhibition of FIE activity can be used to produce larger seeds with increased endosperm. This trait is particularly useful in species in which the endosperm persists in the seed (*e.g.*, monocots, particularly grains).

One of skill will recognize that a number of methods can be used to modulate FIE activity or gene expression. FIE activity can be modulated in the plant cell at the gene, transcriptional, posttranscriptional, translational, or posttranslational, levels. Techniques for modulating FIE activity at each of these levels are generally well known to one of skill and are discussed briefly below.

Methods for introducing genetic mutations into plant genes are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, for example, X-rays or gamma rays can be used.

Alternatively, homologous recombination can be used to induce targeted gene disruptions by specifically deleting or altering the *FIE* gene *in vivo* (*see, generally*, Grewal and Klar, *Genetics* 146: 1221-1238 (1997) and Xu *et al.*, *Genes Dev.* 10: 2411-2422 (1996)). Homologous recombination has been demonstrated in plants (Puchta *et al.*, *Experientia* 50: 277-284 (1994), Swoboda *et al.*, *EMBO J.* 13: 484-489 (1994); Offringa *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 7346-7350 (1993); and Kempin *et al. Nature* 389:802-803 (1997)).

In applying homologous recombination technology to the genes of the invention, mutations in selected portions of an *FIE* gene sequences (including 5' upstream, 3' downstream, and intragenic regions) such as those disclosed here are made *in vitro* and then introduced into the desired plant using standard techniques. Since the efficiency of homologous recombination is known to be dependent on the vectors used, use of dicistronic gene targeting vectors as described by Mountford *et al. Proc. Natl. Acad. Sci. USA* 91: 4303-4307 (1994); and Vaulont *et al. Transgenic Res.* 4: 247-255 (1995) are conveniently used to increase the efficiency of selecting for altered *FIE* gene expression in transgenic plants. The mutated gene will interact with the target wild-type gene in such a way that homologous recombination and targeted replacement of the wild-type gene will occur in transgenic plant cells, resulting in suppression of FIE activity.

Alternatively, oligonucleotides composed of a contiguous stretch of RNA and DNA residues in a duplex conformation with double hairpin caps on the ends can be used. The RNA/DNA sequence is designed to align with the sequence of the target *FIE* gene and to contain the desired nucleotide change. Introduction of the chimeric oligonucleotide on an extrachromosomal T-DNA plasmid results in efficient and specific FIE gene conversion directed by chimeric molecules in a small number of transformed plant cells. This method is described in Cole-Strauss *et al. Science* 273:1386-1389 (1996) and Yoon *et al. Proc. Natl. Acad. Sci. USA* 93: 2071-2076 (1996).

Gene expression can be inactivated using recombinant DNA techniques by transforming plant cells with constructs comprising transposons or T-DNA sequences. *FIE*

mutants prepared by these methods are identified according to standard techniques. For instance, mutants can be detected by PCR or by detecting the presence or absence of *FIE* mRNA, *e.g.*, by Northern blots. Mutants can also be selected by assaying for development of endosperm in the absence of fertilization.

The isolated nucleic acid sequences prepared as described herein, can also be used in a number of techniques to control endogenous *FIE* gene expression at various levels. Subsequences from the sequences disclosed here can be used to control, transcription, RNA accumulation, translation, and the like.

A number of methods can be used to inhibit gene expression in plants. For instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense suppression can act at all levels of gene regulation including suppression of RNA translation (*see*, Bourque *Plant Sci. (Limerick)* 105: 125-149 (1995); Pantopoulos In Progress in Nucleic Acid Research and Molecular Biology, Vol. 48. Cohn, W. E. and K. Moldave (Ed.). Academic Press, Inc.: San Diego, California, USA; London, England, UK. p. 181-238; Heiser *et al. Plant Sci. (Shannon)* 127: 61-69 (1997)) and by preventing the accumulation of mRNA which encodes the protein of interest, *(see,* Baulcombe *Plant Mol. Bio.* 32:79-88 (1996); Prins and Goldbach *Arch. Virol.* 141: 2259-2276 (1996); Metzlaff *et al. Cell* 88: 845-854 (1997), Sheehy et al., *Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt et al., U.S. Patent No. 4,801,340).

The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous *FIE* gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other genes within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though

a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of about 500 to about 1700 nucleotides is especially preferred.

A number of gene regions can be targeted to suppress *FIE* gene expression. The targets can include, for instance, the coding regions, introns, sequences from exon/intron junctions, 5' or 3' untranslated regions, and the like. In some embodiments, the constructs can be designed to eliminate the ability of regulatory proteins to bind to *FIE* gene sequences that are required for its cell- and/or tissue-specific expression. Such transcriptional regulatory sequences can be located either 5'-, 3'-, or within the coding region of the gene and can be either promote (positive regulatory element) or repress (negative regulatory element) gene transcription. These sequences can be identified using standard deletion analysis, well known to those of skill in the art. Once the sequences are identified, an antisense construct targeting these sequences is introduced into plants to control gene transcription in particular tissue, for instance, in developing ovules and/or seed.

Oligonucleotide-based triple-helix formation can be used to disrupt FIE gene expression. Triplex DNA can inhibit DNA transcription and replication, generate site-specific mutations, cleave DNA, and induce homologous recombination (*see, e.g.*, Havre and Glazer *J. Virology* 67:7324-7331 (1993); Scanlon *et al. FASEB J.* 9:1288-1296 (1995); Giovannangeli *et al. Biochemistry* 35:10539-10548 (1996); Chan and Glazer *J. Mol. Medicine (Berlin)* 75: 267-282 (1997)). Triple helix DNAs can be used to target the same sequences identified for antisense regulation.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of *FIE* genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. Thus, ribozymes can be used to target the same sequences identified for antisense regulation.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and

the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Zhao and Pick *Nature* 365:448-451 (1993); Eastham and Ahlering *J. Urology* 156:1186-1188 (1996); Sokol and Murray *Transgenic Res.* 5:363-371 (1996); Sun *et al. Mol. Biotechnology* 7:241-251 (1997); and Haseloff *et al. Nature*, 334:585-591 (1988).

Another method of suppression is sense cosuppression. Introduction of nucleic acid configured in the sense orientation has been recently shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes (*see,* Assaad *et al. Plant Mol. Bio.* 22: 1067-1085 (1993); Flavell *Proc. Natl. Acad. Sci. USA* 91: 3490-3496 (1994); Stam *et al. Annals Bot.* 79: 3-12 (1997); Napoli et al., *The Plant Cell* 2:279-289 (1990); and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184).

The suppressive effect may occur where the introduced sequence contains no coding sequence *per se*, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

For sense suppression, the introduced sequence, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants which are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used. In addition, the same gene regions noted for antisense regulation can be targetted using cosuppression technologies.

Alternatively, *FIE* activity may be modulated by eliminating the proteins that are required for *FIE* cell-specific gene expression. Thus, expression of regulatory proteins

and/or the sequences that control *FIE* gene expression can be modulated using the methods described here.

Another method is use of engineered tRNA suppression of *FIE* mRNA translation. This method involves the use of suppressor tRNAs to transactivate target genes containing premature stop codons (*see*, Betzner *et al. Plant J.*11:587-595 (1997); and Choisne *et al. Plant J.*11: 597-604 (1997). A plant line containing a constitutively expressed *FIE* gene that contains an amber stop codon is first created. Multiple lines of plants, each containing tRNA suppressor gene constructs under the direction of cell-type specific promoters are also generated. The tRNA gene construct is then crossed into the *FIE* line to activate *FIE* activity in a targeted manner. These tRNA suppressor lines could also be used to target the expression of any type of gene to the same cell or tissue types.

As noted above, FIE proteins as products of polycomb group genes are believed to form large complexes *in vivo*. Thus, production of dominant-negative forms of *FIE* polypeptides that are defective in their abilities to bind to other polycomb group proteins is a convenient means to inhibit endogenous FIE activity. This approach involves transformation of plants with constructs encoding mutant FIE polypeptides that form defective complexes with endogenous polycomb group proteins and thereby prevent the complex from forming properly. The mutant polypeptide may vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain. Use of dominant negative mutants to inactivate target genes is described in Mizukami *et al. Plant Cell* 8:831-845 (1996).

Another strategy to affect the ability of an FIE protein to interact with itself or with other proteins involves the use of antibodies specific to FIE. In this method cell-specific expression of FIE-specific Abs is used inactivate functional domains through antibody:antigen recognition (*see*, Hupp *et al. Cell* 83:237-245 (1995)).

Use of nucleic acids of the invention to enhance *FIE* gene expression

Isolated sequences prepared as described herein can also be used to introduce expression of a particular *FIE* nucleic acid to enhance or increase endogenous gene expression. For instance, polycomb genes are known to control cell cycling. Enhanced expression can therefore be used to control plant morphology by controlling whether or not cell division takes place in desired tissues or cells. Enhanced expression can also be used, for instance, to increase vegetative growth by preventing the plant from setting seed. Where overexpression of a gene is desired, the desired gene from a different species may be used to decrease potential sense suppression effects.

One of skill will recognize that the polypeptides encoded by the genes of the invention, like other proteins, have different domains which perform different functions. Thus, the gene sequences need not be full length, so long as the desired functional domain of the protein is expressed.

Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described in detail, below. For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

Preparation of recombinant vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, for example, Weising et al. *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for overexpression, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive

promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumafaciens*, and other transcription initiation regions from various plant genes known to those of skill. Such genes include for example, *ACT11* from *Arabidopsis* (Huang *et al. Plant Mol. Biol.* 33:125-139 (1996)), *Cat3* from *Arabidopsis* (GenBank No. U43147, Zhong *et al.*, *Mol. Gen. Genet.* 251:196-203 (1996)), the gene encoding stearoyl-acyl carrier protein desaturase from *Brassica napus* (Genbank No. X74782, Solocombe *et al. Plant Physiol.* 104:1167-1176 (1994)), *GPc1* from maize (GenBank No. X15596, Martinez *et al. J. Mol. Biol* 208:551-565 (1989)), and *Gpc2* from maize (GenBank No. U45855, Manjunath *et al.*, *Plant Mol. Biol.* 33:97-112 (1997)).

Alternatively, the plant promoter may direct expression of the *FIE* nucleic acid in a specific tissue or may be otherwise under more precise environmental or developmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light. Such promoters are referred to here as "inducible" or "tissue-specific" promoters. One of skill will recognize that a tissue-specific promoter may drive expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue, but may also lead to some expression in other tissues as well.

Examples of promoters under developmental control include promoters that initiate transcription only (or primarily only) in certain tissues, such as fruit, seeds, or flowers. Promoters that direct expression of nucleic acids in ovules, flowers or seeds are particularly useful in the present invention. As used herein a seed-specific promoter is one which directs expression in seed tissues, such promoters may be, for example, ovule-specific (which includes promoters which direct expression in maternal tissues or the female gametophyte, such as egg cells or the central cell), embryo-specific, endosperm-specific, integument-specific, seed coat-specific, or some combination thereof. Examples include a promoter from the ovule-specific *BEL1* gene described in Reiser *et al. Cell* 83:735-742 (1995) (GenBank No. U39944). Other suitable seed specific promoters are derived from the following genes: *MAC1* from maize (Sheridan *et al. Genetics* 142:1009-1020 (1996), *Cat3* from maize (GenBank No. L05934, Abler *et al. Plant Mol. Biol.* 22:10131-1038 (1993), the gene encoding oleosin 18kD from maize (GenBank No. J05212, Lee *et al. Plant Mol. Biol.* 26:1981-1987 (1994)), vivparous-1 from *Arabidopsis* (Genbank No. U93215), the gene

encoding oleosin from *Arabidopsis* (Genbank No. Z17657), Atmyc1 from *Arabidopsis* (Urao *et al.Plant Mol. Biol.* 32:571-576 (1996), the 2s seed storage protein gene family from *Arabidopsis* (Conceicao *et al. Plant* 5:493-505 (1994)) the gene encoding oleosin 20kD from *Brassica napus* (GenBank No. M63985), *napA* from *Brassica napus* (GenBank No. J02798, Josefsson *et al. JBL* 26:12196-1301 (1987), the napin gene family from *Brassica napus* (Sjodahl *et al. Planta* 197:264-271 (1995), the gene encoding the 2S storage protein from *Brassica napus* (Dasgupta *et al. Gene* 133:301-302 (1993)), the genes encoding oleosin A (Genbank No. U09118) and oleosin B (Genbank No. U09119) from soybean and the gene encoding low molecular weight sulphur rich protein from soybean (Choi *et al. Mol Gen, Genet.* 246:266-268 (1995)).

In addition, the promoter sequences from the *FIE* genes disclosed here can be used to drive expression of the *FIE* polynucleotides of the invention or heterologous sequences. The sequences of the promoters are identified below.

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences (*e.g.*, promoters or coding regions) from genes of the invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

Production of transgenic plants

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl. Acad. Sci. USA* 82:5824

(1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:70-73 (1987).

Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch et al. *Science* 233:496-498 (1984), and Fraley et al. *Proc. Natl. Acad. Sci. USA* 80:4803 (1983).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype such as increased seed mass. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176, MacMillilan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium, Arachis, Asparagus, Atropa, Avena, Brassica, Citrus, Citrullus, Capsicum, Carthamus, Cocos, Coffea, Cucumis, Cucurbita, Daucus, Elaeis, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoscyamus, Lactuca, Linum, Lolium, Lupinus, Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Olea, Oryza, Panieum, Pannesetum, Persea, Phaseolus, Pistachia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna,* and *Zea*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into

other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

Seed obtained from plants of the present invention can be analyzed according to well known procedures to identify plants with the desired trait. If antisense or other techniques are used to control Fie gene expression, Northern blot analysis can be used to screen for desired plants. In addition, the presence of fertilization independent reproductive development can be detected. Plants can be screened, for instance, for the ability to form embryo-less seed, form seed that abort after fertilization, or set fruit in the absence of fertilization. These procedures will depend, part on the particular plant species being used, but will be carried out according to methods well known to those of skill.

The following Examples are offered by way of illustration, not limitation.

Example 1

The following example describes methods used to identify the *fie* mutants. The methods described here are generally as described in Ohad *et al.*, *Proc. Natl. Acad. Sci. USA* 93:5319-5324 (1996).

MATERIALS AND METHODS

Growth and Phenotype of Plants. Plants were grown under low humidity conditions (less than 50%) in glass houses under 16 hr light/8 hr dark photoperiods generated by supplemental lighting. Plants were grown at high humidity (greater than 80%) in a lighted incubator (Percival, Boone, Iowa).

To test for fertilization-independent development, flower buds from plants that had not yet begun to shed pollen (stage 12; (Smyth, D. R., *et al.*, *Plant Cell* 2: 755-767 (1990))) were opened, immature anthers were removed, and the flower bud was covered with a plastic bag. Seven days later, the silique was measured, dissected, and the number of seed-like structures and degenerating ovules were counted. To determine the frequency of seed abortion following fertilization, siliques were harvested 10 days after self-pollination, dissected, and wild-type and aborted seeds were counted.

Genetic Mapping. Heterozygous FIE/fie (Landsberg erecta ecotype) plants were crossed as males with female plants (Columbia ecotype). Because the mutant fie allele is only transmitted through the male gametophyte, FIE/fie progeny were crossed as males a second

time to female gl1/gl1 (Columbia ecotype) plants. Approximately fifty-five progeny were scored for the segregation of the wild-type FIE and mutant fie alleles and for alleles of molecular markers as described previously (Bell, C., *et al.*, *Genomics* 19: 137-144 (1994)). This analysis indicated that *fie3* is located at approximately position 30 on chromosome three, *fie2* is located at approximately position 65 on chromosome two, and *fie1* is located at approximately position 2 on chromosome one. Genetic recombination frequencies and map distances were calculated according to Koornneef and Stam (Koornneef, M., *et al.*, *Methods in Arabidopsis Research*, pp. 83-99 (1992)) and Kosambi (Kosambi, *Ann. Eugen.*, 12: 172-175 (1944)).

Light Microscopy. Nomarski photographs of whole-mount embryos and endosperm were obtained by fixing longitudinally slit siliques in an ethanol:acetic acid (9:1) solution overnight, followed by two washes in 90% and 70% ethanol, respectively. Siliques were cleared with a chloral hydrate:glycerol:water solution (8:1:2, w:v:v) (Berleth, T., *et al.*, *Devel* 118: 575-587 (1993)). Whole mount preparations were fixed and stained with hematoxylin (Beeckman, T., *et al.*, *Plant Mol Biol Rep* 12: 37-42 (1994)). Embryo and endosperm were photographed with a Zeiss Axioskop microscope (Carl Zeiss, Inc., Oberkochen, Germany) using Nomarski optics that permits visualization of optical sections within the seed.

GUS Histochemical Assays. GUS activity was detected histochemically as described previously by (Beeckman, T., *et al.*, *Plant Mol Biol Rep* 12: 37-42 (1994)).

Image Processing. Photographs were scanned using a Microtek scanner. Pictures were processed for publication using Adobe Photoshop 3.0 and printed on a Tektronix Phaser 400 color printer.

RESULTS

Isolation of Mutant Lines. To begin to understand mechanisms that initiate reproductive development, we generated mutant Arabidopsis plants that undergo several reproductive processes in the absence of fertilization. Arabidopsis plants homozygous for the conditional male sterile pop1 mutation (Preuss, D., *et al.*, *Genes and Devel* 7: 974-985 (1993)) were used as the parental strain (Landsberg erecta ecotype). Fertility in pop1 plants is sensitive to humidity because pop1 pollen do not hydrate properly due to a defect in wax biosynthesis.

When grown at permissive condition, high relative humidity (>80%), pop1 plants were male fertile and produced long siliques with many viable seeds. By contrast, when grown at non-permissive condition, low relative humidity (<50%), pop1 plants were male sterile and produced short siliques with no seeds. Thus, silique elongation is a marker for reproductive events. To isolate mutations, homozygous pop1 seeds were mutagenized with ethylmethansulfonate (EMS) and approximately 50,000 M1 plants were screened for silique elongation at non-permissive conditions. Rare M1 plants were identified that displayed heterozygous sectors with elongated siliques. These plants were transferred to permissive conditions to insure the production of viable M2 seed. Plants from M2 and M3 families grown at non-permissive conditions were rechecked for non-sectored silique elongation. To eliminate any effects of the pop1 mutation, or other EMS-induced lesions on the mutant phenotype, mutant plants were backrossed twice, as males, to wild-type plants. After removing the pop1 mutation, fertilization-independent phenotypes were confirmed after manual removal of anthers from immature flowers before pollen was shed. A total of twelve lines were identified that displayed elongated siliques in the absence of fertilization.

Fertilization-Independent Endosperm, Seed Coat and Silique Development. In a representative line chosen for further study, heterozygous plants produced by back crosses to wild-type plants generated elongated siliques after anther removal with numerous seed-like structures. These results indicated that heterozygous mutant plants were capable of silique elongation and seed-like structure development in the absence of fertilization. We compared the development of the mutant seed-like structures to that of wild-type seeds. After fertilization, the endosperm nucleus replicated and daughter nuclei migrated into the expanding central cell. Ultimately, a syncytium of endosperm nuclei was produced. Nuclear divisions of the endosperm preceded the zygotic divisions that formed the globular stage embryo. Embryo, endosperm or seed coat development did not occur in wild-type plants in the absence of fertilization. Development of the ovule and female gametophyte in heterozygous mutant plants was normal. Just prior to flower opening, female gametophytes in these plants contained a single, prominent central cell nucleus. Subsequently, in the absence of fertilization, central cells with two large nuclei were detected. Further divisions resulted in the production of additional nuclei that migrated into the expanded central cell. Later in development, a nuclear syncytium was formed with abundant endosperm nuclei. These results indicated that the central cell in mutant female gametophytes initiated

endosperm development in the absence of fertilization. We have named this mutation fie for fertilization-independent endosperm. By contrast, replication of other nuclei in fie female gametophytes (egg, synergid, or antipodal) was not detected. Thus, the fie mutation specifically affects replication of the central cell nucleus.

We analyzed the frequency of multinucleate central cell formation in fie female gametophytes by comparing the percentage of multinucleate central cells at three, five, and six days after emasculation of heterozygous FIE/fie and control wild-type flowers. At each time point, only 3% to 5% of wild-type central cells had more than one nucleus. Because none had more than two nuclei, most likely, these represented central cells with haploid nuclei that had not fused during female gametophyte development. By contrast, the percentage of central cells in female gametophytes from FIE/fie siliques with two or more nuclei increased from 21% to 47% over the same time period. These results indicated that the fie mutation caused a significant increase in formation of multinucleate central cells in the absence of fertilization. The fact that close to 50% of the female gametophytes in heterozygous plants had multinucleate central cells suggested that fie is a gametophytic mutation because a 1:1 segregation of wild-type and mutant fie alleles occurs during meiosis.

We compared the fertilization-independent development of the maternal seed coat in FIE/fie seed-like structures to that of fertilized wild-type seeds. The seed coat in wild-type Arabidopsis is generated by the integuments of the ovule and surrounds the developing embryo and endosperm. Similarly, FIE/fie ovule integuments formed a seed coat that surrounded the developing mutant endosperm. These results indicated that the fie mutation activated both endosperm development and maternal sporophytic seed coat and silique differentiation that support reproduction. No other effects on sporophytic growth and development were detected in FIE/fie plants.

The fie3 Mutant Allele Is Not Transmitted by the Female Gametophyte to the Next Generation. To understand the mode of inheritance of the fie mutation, we analyzed the progeny of reciprocal crosses. FIE3/fie3 females, crossed to wild-type males, produced siliques with approximately equal numbers of viable seeds with normal green embryos and nonviable white seeds with embryos aborted at the heart stage (344:375, 1:1, $c2 = 1.3$, $P > 0.2$). Viable seeds from this cross were germinated and all 120 F1 progeny generated were wild-type. That is, none of the F1 progeny had significant levels of F2 aborted seeds in their siliques after self-pollination. Nor did the F1 progeny demonstrate fertilization-independent

development. This indicated that presence of the fie mutant allele in the female gametophyte, even when the male provided a wild-type allele, resulted in embryo abortion. Thus, the fie mutation is not transmitted by the female gametophyte to the next generation. To study transmission of fie through the male gametophyte, we pollinated female wild-type plants with pollen from male FIE3/fie3 plants. Siliques from these crosses contained no aborted F1 seed. F1 plants were examined and a 1:1 segregation of wild-type and FIE3/fie3 genotype was observed (62:58, $c2 = 0.13$, $P > 0.5$). This indicated that wild-type and mutant fie3 alleles were transmitted by the male gametophyte with equal efficiency. That is, fie does not affect male gametophyte, or pollen grain, function. Results from reciprocal crosses were verified by analyzing the progeny from self-pollinated FIE3/fie3 plants. Self-pollinated siliques displayed 1:1 segregation of normal and aborted seeds (282:286, $c2 = 0.03$, $P > 0.8$). Viable seed from self-pollinated siliques were germinated and a 1:1 (71:64, $c2 = 0.36$, $P > 0.5$) segregation of wild-type and FIE3/fie3 progeny was observed. These results confirmed that inheritance of a fie mutant allele by the female gametophyte resulted in embryo abortion, and that inheritance of a fie mutant allele by the male gametophyte did not affect pollen function. Thus, the wild-type FIE3 allele probably carries out a function unique to the female gametophyte and does not appear to be needed for male fertility.

In contrast, *fie1* and *fie2* mutant alleles were transmitted at low frequencies (about 1% of normal) through the female gametophyte. In this way, *fie1* homozygous mutants and *fie2* homozygous mutants were obtained that appeared to display normal vegetative growth and development.

DISCUSSION

In wild-type plants, fertilization initiates embryogenesis and endosperm formation, and activates maternal seed coat and silique development. The results presented here indicate that specific aspects of plant reproductive development can occur in FIE/fie plants in the absence of fertilization. These include silique elongation, seed coat formation, and endosperm development. Morphological analysis shows that early aspects of fertilization-independent fie endosperm development closely resemble fertilized wild-type endosperm development. First, the fie central cell nucleus is stimulated to undergo replication. Second, nuclei that are produced migrate from the micropylar end of the central cell and take up new positions in the central cell. Third, the developing fie central cell expands to form an endosperm cavity. Thus, the requirement for fertilization to initiate these

early events in endosperm formation has been eliminated by the fie mutation. This suggests that FIE plays a role in a signal transduction pathway that links fertilization with the onset of central cell nuclear replication and early endosperm development.

Mechanisms for Regulation of Endosperm Development by FIE. One can envision two possible mechanisms for how FIE regulates replication of the central cell nucleus in response to fertilization. The protein encoded by the FIE gene may be involved in a positive regulatory interaction. In this model, FIE is required for the central cell to initiate endosperm development. Normally, fertilization is needed for the presence of active FIE protein. The fie mutation results in the presence of active protein in the absence of fertilization. Alternatively, FIE may by involved in a negative regulatory interaction. In this model, the function of FIE protein is to prevent the central cell from initiating endosperm development, and fertilization results in the inactivation of FIE protein. The fie mutation results in the production of inactive protein, so that fertilization is no longer required to initiate endosperm development. However, complementation experiments using transgenic plants indicate that *FIE1* and *FIE3* alleles are dominant over their respective mutant alleles. This indicates that the wild-type allele is involved in a negative regulatory interaction. Recently, it has been shown that cyclin-dependent kinase complexes, related to those that function in mammals, control the induction of DNA synthesis and mitosis in maize endosperm (Grafi, G.*et al.*, *Science* 269: 1262-1264 (1995)). Because fie stimulates replication of the central cell, fie may, either directly or indirectly, impinge upon cell cycle control of the central cell nucleus, allowing replication to take place in the absence of fertilization.

Communication between the fie Female Gametophyte and the Sporophytic Ovule and Carpels. The analysis of FIE/fie mutant plants has provided clues about interactions between endosperm and maternal sporophytic tissues. FIE/fie ovule integuments surrounding a mutant fie female gametophyte initiate seed coat development, whereas FIE/fie integuments in contact with a quiescent wild-type female gametophyte do not develop. This suggests that the FIE/fie ovule integuments initiate seed coat differentiation in response to a signal produced by the fie female gametophyte. We propose that the source of the signal is the mutant fie central cell that has initiated endosperm development, although

we cannot rule out the participation of other cells in the fie female gametophyte. In wild-type plants, most likely, fertilization of the central cell produces an endosperm that activates seed coat development. This is consistent with experiments showing that the maize endosperm interacts with nearby maternal cells (Miller, M. E., *et al.*, *Plant Cell* 4: 297-305 (1992)). FIE/fie plants also display fertilization-independent elongation of the ovary to form the silique. We propose that a signal is produced by the developing seed-like structures to initiate silique elongation. This is in agreement with experiments suggesting that seeds are the source of hormones, auxins and gibberellins, that activate fruit development (Lee, T. D. *Plant Reproductive Ecology*, pp. 179-202 (1988)). Taken together, these results suggest that the fertilized female gametophyte activates maternal developmental programs.

Relationship between fie and Apomixis. Certain plant species display aspects of fertilization-independent reproductive development, including apomictic generation of embryo and endosperm, and development of the maternal seed coat and fruit (reviewed in (Koltunow, a. *Plant Cell* 5: 1425-1437 (1993)). The fie mutation reveals that Arabidopsis, a sexually reproducing plant, has the genetic potential for aspects of fertilization-independent reproductive development. It is not known whether the mechanism of fertilization-independent endosperm development conferred by the fie mutation is the same as autonomous endosperm formation observed in certain apomictic plant species. However, the fact that the fie phenotype is caused by a single genetic locus substantiates the view that the number of genetic differences between sexually and asexually reproducing plants is small (Koltunow, a. M., *et al.*, *Plant Physiol* 108:1345-1352 (1995)).

Example 2

This example describes cloning of two *Fie* genes, *Fie1* and *Fie3*.

Cloning The *FIE3* Gene.

a. Mapping the position of the *fie3* gene genetically. The *fie3* mutation was initially mapped to position 30 on chromosome 3, between *AXR2* (auxin resistant dwarf) and *EMB29* (embryo lethal). Next, two sets of F2 plants with recombination breakpoints in the *fie3* gene region were obtained. One set was between *emb29* and *fie3* and the other set was between *axr2* and *fie3*. As shown in Figure 1A, these recombinants were used to map the *fie3* gene relative to molecular markers (NDR, CH18, CH18S, BO20, AG20, KN1 and

E13F12) that were obtained from overlapping YAC (yUP13F12), BAC (T1B4 and T4N1) and cosmid clones (Fig. 1A). YAC and BAC clones were obtained from the Arabidopsis Stock Center (Ohio State University, USA). Cosmid subclones were generated in my laboratory. As shown in Fig. 1A and 1B, this genetic analysis indicates that the *fie3* gene resides within the 25 Kb region between the BO20 and AG20 markers.

b. Mapping the position of the *fie3* gene by complementation experiments. To more precisely localize the *fie3* gene, we analyzed a series of overlapping cosmid clones (BO20, GM15, AG20 and EI12) that span the *fie3* gene region. Each cosmid clone was tested for its ability to complement the *fie3* mutation in transgenic plants. Only cosmid GM15 complemented the *fie3* mutation (Fig. 1A). These results indicate that an essential portion of the *fie3* gene is in the 10 Kb region that is unique to cosmid GM15. As shown in Fig. 1B, we have cloned DNA that spans this essential portion of the *fie3* gene and have determined its DNA sequence. As shown in Fig. 2, analysis of the sequence using the GENSCANW program revealed a gene with an open reading frame. The predicted cDNA sequence and predicted amino acid sequence are shown in SEQ ID NO:3 and SEQ ID NO:4, respectively. Comparing the predicted amino acid sequence to those in public data bases revealed significant homology to the WD40 family of Polycomb Group genes, and in particular, the "extra sex combs" gene in *Drosophila*. Figure 3 shows the position of primers used to PCR amplify this region. SEQ ID NO:5 provides the genomic DNA sequence of the WD40/Polycomb gene, plus approximately 3.8 Kb of 5'-flanking sequences and 0.3 Kb of 3'-flanking sequences, plus the sequence of primers used to PCR amplify this region. The transcription start site in SEQ ID NO:5 is at position 3,872. Thus, the promoter sequence for *FIE3* is located between position 1 and 3,872. The 5'-flanking and 3'-flanking regions contain regulatory DNA sequences that control the expression of this gene.

Cloning the *FIE1* Gene.

a. Mapping the position of the *FIE1* gene genetically. The *fie1* mutation was initially mapped to position 3 on chromosome 1, between *AXR3* (auxin resistant dwarf) and EMB60 (embryo lethal). Next, two sets of F2 plants with recombination breakpoints in the *FIE1* gene region were obtained. One set was between emb60 and *fie3* and the other set was between *axr3* and *fie3*. These recombinants were used to map the *fie3* gene relative to

molecular markers (Fig. 4) that were obtained from an overlapping series of YAC and BAC clones from the Arabidopsis Stock Center (Ohio State University, USA).

b. Mapping the position of the *FIE1* gene by complementation experiments. To more precisely localize the *FIE1* gene, a series of overlapping cosmid clones (2-9, 6-22, 2-8) that span the *FIE1* gene region were analyzed (Fig. 4). Each cosmid clone was tested for its ability to complement the *fie1* mutation in transgenic plants. Only cosmid 6-22 complemented the *fie1* mutation. The cosmids were analyzed for genes with open reading frames. Figure 5 shows that a single gene was present on the complementing cosmid (6-22) that was not fully encoded on either of the non-complementing cosmids (2-9 and 2-8). By RTPCR and 5'-race, the cDNA sequence of this gene and predicted amino acid of its protein were obtained (SEQ ID NO:1 and SEQ ID NO:2, respectively). Comparison of the predicted amino acid sequence to those in public data bases revealed significant homology to the SET family of Polycomb Group Genes (e.g., Enhancer of Zeste in *Drosophila* and Curly Leaf in *Arabiopsis*). We compared the wild-type and *fie1* mutant sequence in 6-22. The only difference is a single base pair change that creates a premature translation stop codon in the 5'-end of the set/polycomb group gene. The base pair change is at position 823 (C -> T) on the cDNA sequence shown in SEQ ID NO:1.

SEQ ID NO:6 shows the genomic sequence of the *FIE1* SET/polycomb gene, plus approximately 2 Kb of 5'-flanking sequences and approximately 0.7 Kb of 3'-flanking sequences. The translation start site is located at position 2036 of SEQ ID NO:6. Thus, the promoter sequence is located between position 1 and position 2036.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference.